DOCUMENT RESUME

ED 458 271 TM 033 434

AUTHOR Harwell, Michael
TITLE Future Directions and Suggestions for Improving the Use of Statistical Methods in Educational Research.
PUB DATE 2001-04-00
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Quality; *Educational Research; *Research Methodology; Scholarly Journals; *Statistical Analysis; Writing for Publication

ABSTRACT

This discussion of the use of statistical methods in educational research considers the frequent misuse of statistical procedures, the factors responsible for the misuse, and the remedies that might reduce the misuse of statistical procedures. There are longstanding concerns about the quality of statistical work in general as well as in educational research, but there is little formal documentation of the misuse of statistical methods by graduate students. Anecdotal evidence suggests that the misuse is not confined to students, but can be found in the work of faculty members and researchers. Misuse of statistical procedures by students is in large part due to inadequate instruction, sometimes resulting from poorly prepared teachers or inadequate texts, or even an incoherent curriculum. The magnitude of misuses of statistical procedures by researchers also suggests inadequate statistical training. Any remedies for these problems will need to be tailored to various types of students and educational researchers. Some possible solutions are: (1) put education back into educational statistics; (2) make instructors accountable for their teaching; (3) insist on adequate training for students; (4) include educational statisticians as reviewers for scholarly journals; (5) value research in statistical education; and (6) establish a professional "home" for educational statisticians. (Contains 35 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 458 271

Future Directions and Suggestions for Improving The Use of Statistical Methods in Educational Research

Michael Harwell

University of Minnesota

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM033434

M. Harwell

1

BEST COPY AVAILABLE

Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
Correspondence concerning this paper should be directed to Michael Harwell, Department of Educational Psychology, 323 Burton Hall, 178 Pillsbury Drive SE, University of Minnesota, Minneapolis, MN 55455 (email: harwe001@umn.edu)

Educational statisticians working in academic settings play a key role in the development, application, and teaching of statistical procedures. They are also responsible for ensuring, to the extent possible, that statistical procedures are used properly by students, faculty colleagues, and educational researchers in general. Unfortunately, there is evidence that these procedures are frequently misused. The misuse ranges from the obvious, such as performing ANOVA F-tests on dichotomous data, to the subtle, for example, problems of compounded Type I error rates. The misuse occurs across a spectrum of statistical procedures, although it's likely that complex procedures like MANOVA, hierarchial linear modeling (HLM), and structural equation modeling (SEM) are particularly prone to misuse.

The discussion of future directions and some suggestions for improving the use of statistical methods in educational research is organized around three questions (1) How severe is the misuse of statistical procedures in education by students and researchers? (2) What factors are responsible for the misuse? (3) What remedies are needed to reduce the misuse of statistical procedures in education? The questions and the proposed remedies are not new; however, the continuing misuse of statistical procedures in educational research suggests that there is still a need to raise these issues and to revisit remedies that could reduce misuse.

How Severe is the Misuse of Statistical Procedures in Educational Research?

Clearly, there are longstanding concerns about the quality of statistical work in general (Berkson, 1942; Hogg, 1990, 1991; Hooke, 1980; Snee, 1993), as well as in educational research (Brewer, 1985; Keselman, Huberty, Lix, Olejnik, Cribbie, Donohue, Kowalchuk, Lowman, Petosky, Keselman, & Levin, 1998; Lykken, 1968; Wilson, 1973). The question is not whether statistical procedures are misused, but the severity of the misuse and its consequences on inferences.

There is little formal documentation of the misuse of statistical methods by graduate students in schools and colleges of education. Curtis and Harwell (1998) reported the results of a survey of educational statistics faculty who responded to questions about the perceived statistical competence of students. Among the findings were that 36% of the respondents thought that > 50% of the students with whom they had contact with in the past five years could not perform ANOVA competently, and 59% thought that > 50%

were not competent in ordinary least squares multiple regression. Virtually all of the respondents thought that > 50% of the students could have benefited from additional statistics coursework, 44% thought that > 50% could have benefited from one or two additional courses, and 43% thought that at least half could have benefited from three or more additional statistics courses. These results support the notion that graduate students frequently do not receive adequate statistical training, increasing the likelihood of misusing statistical procedures.

Along the same lines, anecdotal evidence that statistical methods are frequently misused by students appears to be widely shared by educational statisticians in faculty roles. This perception no doubt comes from first-hand experience in teaching, service on student thesis/dissertation committees, and advising students on statistical issues. Most educational statisticians can probably describe their experiences with the misuse of statistical methods by students. My favorite (worst) example involved a doctoral student who used a survey to collect data and discovered that more than 70% of the data values were missing. After the missing data issue was raised at the dissertation proposal meeting (possible biasing effects, reduced precision of estimation, etc.), I was surprised to receive a draft of the entire dissertation about two months later, and even more surprised to learn that there were no missing data. The student, on the advice of their advisor, had used a BMDP program to impute values and eliminate the missing data problem without any reference to the assumptions necessary to justify this practice (e.g., missing data should be missing at random).

This example suggests that misuse of statistical procedures is not limited to students. Certainly, attendance at professional conferences like AERA contributes to the view among educational statisticians that statistical procedures are frequently misused. More formally, several studies have documented the misuse of statistical procedures by our colleagues. For example, Keselman, et al., (1998) performed an extensive literature review that provided evidence of the misuse of frequently employed statistical tools (ANOVA, MANOVA, ANCOVA) in published articles. Harwell and Gatti (in press) surveyed papers in three prominent journals in educational research (American Educational Research Journal, Journal of Educational Psychology, Sociology of Education) in 1997 for the nature of the variables used in normal-theory tests. Overall, 76% of the more than 700 variables assumed to be normally-distributed possessed a

Likert –type scale and could not possibly be normally-distributed. Almost all of these studies employed HLM or SEM, which for non-normally distributed data are known to produce biased estimators and significance tests. Other examples were provided by Glisson and Hudson (1981), who described the misuse of statistical procedures in admission decisions in education, Daniel (1998), who described the general misuse of significance tests, and Helberg (1996), who described several common misuses of statistical procedures by educational researchers.

My favorite (worst) example of misuse came during my graduate training and involved a faculty member with an unnatural fondness for creating new variables out of existing ones. One variable-creation frenzy produced approximately 1000 new variables. At the faculty member's request, approximately 1000 two-sample t-tests were performed in which the same groups of males and females were compared, each at $\alpha = .05$. I was surprised when only 42 of the approximately 1000 t-tests were statistically significant, and shocked when the researcher expressed delight over those results that were significant. My concern that the data had not even behaved according to chance was met with a blank stare.

Readers can no doubt supplement the above examples with their own experiences with the misuse of statistical procedures by students and faculty colleagues. Consequences of this misuse include poorly estimated parameters and larger than acceptable probabilities of rejecting true statistical null hypotheses (Type I errors) and retaining false statistical null hypotheses (Type II errors). It is impossible to accurately gauge the effect of misuse on inferences, but it seems fair to conclude that it is non-negligible. Although largely anecdotal, it appears that many educational statisticians believe that misuse of statistical procedures is a widespread and serious problem.

What Factors Are Responsible for the Misuse of Statistical Procedures by Students?

Misuse of statistical procedures by students is largely an instructional problem, involving failures in classroom instruction as well as failures in teaching outside the classroom (e.g., dissertation work). There are many views of what is wrong in statistics instruction, particularly in introductory classes. Only a few of the factors that are at least partly responsible for the misuse of statistical procedures by students (some of whom continue to misuse statistical procedures as educational researchers) are described below.

It's The Students

Inadequate Training

If anecdotal evidence has any merit, this is a popular view among statistics instructors. Perhaps the simplest explanation for misuse of statistical procedures is that many students simply do not receive enough statistical training (see Curtis & Harwell, 1998). In many schools and colleges of education the minimum number of statistics courses a doctoral student must sit for is two, hardly adequate to prepare a student to use the kinds of statistical procedures they may need in their work (multiple regression, MANOVA, HLM, etc.). Many would agree that inadequate statistical training opens the door to misuse.

Heterogeneity of Students in Statistics Classes

A more complex explanation relates to classroom heterogeneity. Most instructors in introductory statistics classes taught in schools and colleges of education would probably agree that heterogeneity of student backgrounds, preparedness, and motivation poses a major instructional challenge. This view is often seasoned with the notion that each statistics class provides a unique challenge (Brogan & Kutner, 1986), making one-size-fits-all prescriptions for statistics instruction difficult. Moreover, what works well in statistics classes at one university may not work well at another, further complicating matters for prescribing changes in statistics instruction.

Another constituent of classroom heterogeneity is that some students lack the prerequisite skills necessary to succeed in these courses. This problem may be exacerbated when enrollment prerequisites are not rigorous or existing prerequisites are not enforced. Although anecdotal, this view is sometimes accompanied by the claim that today's students are inferior to those from, say, 15-20 years ago (Hogg, 1991, made this claim for students in general). However, the merit of this view is weakened somewhat by those of us who were in graduate school 15-20 years ago and who heard precisely the same comment, that current students simply are not as good as those from 15-20 years ago.

Variation in the motivation of students also may contribute to classroom heterogeneity. Those who are required to sit for 1-2 semesters of introductory statistics as a degree requirement may not be as motivated as students majoring in educational statistics, yet these students may be in the same class.

One way to recognize and perhaps sort out this heterogeneity is to characterize students in educational statistics courses along some meaningful dimension. Elmore and Burney (1997) created four categories to represent interaction patterns between the leadership in a school district (superintendent, deputy superintendent, director of professional development, etc.) and district schools (principal, teachers, support staff). Adapted to an educational statistics setting, the categories of students are:

• *With-the-Drill*

With-the-Drill students are, from the perspective of many statistics instructors, close to the "ideal." These students are highly motivated, adhere to the rules and expectations of the course, are detail-oriented, and show substantial self-discipline. These are the students who religiously show up for office hours and review sessions, read the assigned material before class, spend prodigious amounts of time studying, and are genuinely engaged in the class. They typically have had at least some prior statistics or mathematics training in college. They rarely demonstrate an appetite for independent learning, but can be counted on to fulfill the course requirements and to perform well.

• *Free-Agents*

Free-Agents are students who are typically well-prepared and are independent learners. In many cases they have had previous statistical training. They quickly grasp the material and often value the use of statistical proofs and instructional efforts to link statistical concepts. However, their motivation and self-discipline often lags far behind With-the-Drill students. They do not show up regularly for office hours or review sessions and cannot be counted on to read assigned material before a class. There is frequently little evidence that they spend much time studying, although they typically perform well on exams. These students are treated as "free-agents" in the sense that instructors usually do not worry about them, confident that they will understand the material (often with far less effort than With-the-Drill students) and perform well without much assistance.

• *Watch-List*

Watch-List students are those who, while motivated, are often ill-prepared. They frequently have had no previous statistics or mathematics courses in college, and frequently claim to be phobic about

numbers. As a result, they often struggle in Herculean fashion. They appear to be fully engaged in the course, take advantage of available resources, are sometimes highly motivated, spend prodigious amounts of time studying, and yet still perform poorly. They rarely show an interest in independent learning, and instructors typically spend more time with these students than those in any other category. Watch-list students are similar in many ways to With-the-Drill students, but their typical lack of undergraduate training in statistics or mathematics and their self-proclaimed lack of confidence with numbers set them apart.

• *Off-the-Screen*

These are students who are poorly prepared, undisciplined, and unmotivated. They uniformly fail to take advantage of available resources, rarely showing up for office hours or review sessions. These students fail to adhere to the rules and expectations of the course and show a remarkable lack of self-discipline. They are easy to pick out because they rarely read the assigned material before class, spend miniscule amounts of time studying, and demonstrate no appetite for independent learning. They are genuinely disengaged from the class, yet are often surprised when they perform poorly.

From an instructional perspective, Free Agents and With-the-Drill students offer the fewest instructional challenges, and Watch-List and Off-the-Screen students the most. It's probably fair to say that many introductory statistics classes contain all four types of students, as well as others not represented by the above categorization.

It's The Textbook

Faculty and student complaints about statistics textbooks are the stuff of legends. Even within the educational statistics community there is substantial disagreement over what constitutes a "good" statistics textbook, although most statistics instructors would probably agree that the textbook can play an important role in student learning. Cobb (1987) indicated that the quality of a statistics textbook could be judged by the quality of its exercises. In the same spirit, Singer and Willett (1990) argued that statistics texts should include real datasets as a way of engaging students from varied backgrounds and to demonstrate the proper use of statistical procedures with less than ideal data (e.g., data containing outliers). Following the Cobb and Willett and Singer criteria (among others), Harwell, Herrick, Curtis, Mundfrom, and Gold (1996) evaluated 10

introductory statistics texts used in education and reported substantial heterogeneity among these texts on several dimensions. These results support the view that there is substantial variation in what is believed to constitute a good statistics textbook. The textbook probably matters most to the With-the-Drill and Watch-list students, and less so for Free-Agent students. The effect of the textbook on Off-the-Screen students is hard to gauge.

It's The Instructor

Another explanation for the misuse of statistical methods by students is rooted in the attitudes of some statistics instructors. Consider the research on attitudes toward teaching in K-12 education. Many of the school reform efforts in the U.S. espouse a vigorous theory of educational equity for K-12 education, one that argues that every student can learn under the right conditions (Bryk, Sebring, Kerbow, Rollow, & Easton, 1998; Elmore, & Burney, 1999; Knapp, & McLaughlin, 1999), and that teachers (along with other school personnel) have an obligation to adopt this view. Does this perspective transfer to introductory educational statistics classes? Do statistics instructors believe that all students can learn if given the proper support? Even Watch-List and Off-the-Screen students?

The extent to which some statistics instructors hold the view that some students cannot or will not learn the material is hard to gauge, but it will probably have its greatest effect on Watch-List and Off-the-Screen students because the instruction will likely be oriented toward the Free-Agent and With-the-Drill students. It is easy to see how this could lead to significant misuse of statistical procedures. For example, Watch-List students who were largely ignored in introductory statistics classes may go on to perform statistical analyses for their theses and dissertations. In the absence of high quality statistics instruction targeting their learning needs, the likelihood of misuse of statistical procedures seems high.

It's The Adjunct Instructor

A variation of *It's The Instructor* is the view that using adjunct instructors to teach statistics courses, which appears to be on the increase, can have negative effects on student learning (Keselman, et al., 1998; Cobb, 1993). The usual problem is that these instructors do not receive the support and oversight they may need.

Failure to Adequately Convey the Need For Careful Use of Statistical Procedures

One of the concerns that emerges regularly in the statistics literature is the sloppiness with which many data analyses are conducted. Neter, Kutner, Nachtsheim, and Wasserman (1996, p. 756) described the sequence of developing and using any statistical model as

i. Examine whether the proposed model is appropriate for the set of data at hand.

ii. If the proposed model is not appropriate, consider remedial measures, such as transformation of the data or modification of the model.

iii. After review of the appropriateness of the model and completion of any necessary remedial measures and an evaluation of their effectiveness, inferences based on the model can be undertaken.

Consider a single-factor, fixed-effects ANOVA with dependent variable Y. An appropriate series of steps for performing this analysis is the following:

a. Decide that an ANOVA F-test should be performed.

b. Examine the viability of the underlying linear model. Is a linear model sensible? Is the notion of a constant additive treatment effect credible? (Callaert, 1999) How likely is it that scores are independent of one another?

c. Plot the scores and examine the plot for evidence of nonnormality, heteroscedasticity, outliers, truncation, etc. If present, adjust the analysis accordingly. For example, consider a nonparametric test if the data appear to show a nonnormal distribution, or employ a nonlinear data transformation to try to reduce heteroscedasticity (Draper, 1988).

d. Perform the analysis and reject/retain the hypothesis of equal population means.

e. If the hypothesis is rejected, compute estimates of effect size and perform appropriate post hoc analyses.

f. Carefully draw inferences from the statistical findings, taking into account whether a randomized or nonrandomized design was used.

It is not much of a stretch to suggest that many ANOVAs are performed using only steps a and d, and perhaps e. Clearly, students who perform ANOVA in this way are much more likely to misuse this procedure. Similar behavior when performing more complex procedures, such as ANCOVA, multiple regression, HLM, and SEM substantially increases the likelihood of misuse. Other things being equal, this sort of misuse is most likely with Off-the-Screen and Watch-List students, although it sloppiness may also appeal to Free-Agents.

A plausible explanation for this undesirable practice is that the need for careful and thoughtful use of statistical procedures has not been adequately or consistently conveyed to students in classroom instruction, as well as other instructional settings (e.g., dissertation work). Some educational statisticians might respond that statistical instruction and practice is largely dedicated to training students to correctly complete steps a-f, and that the responsibility for failing to practice what has been taught should be wholly or partly borne by the student (i.e., *It's The Student*).

Inefficient Instructional Methods

One of the more active areas of research on teaching statistics focuses on the way that information is delivered to students. A general finding in this area is that traditional lecture formats may poorly serve students with particular learning styles (Garfield, 1993; Giraud, 1997). There is evidence that this occurs because some instructors adopt a wholistic approach (present the big picture first, then work down to the details), whereas others adopt a serialistic approach (begin with the details and work up toward the big picture) (Pask, 1976). For example, an instructor employing a wholistic approach for students sharing a serialistic learning style may lead to significant learning difficulties.

Other research has consistently supported the use of cooperative learning, small groups, and hands-on" activities in introductory statistics classes (Garfield, 1993, 1995; Giraud, 1997; Magel, 1998). These findings are not surprising given the voluminous literature in K-12 education documenting the potential value of small group and cooperative learning. There is also evidence that the traditional assessment structure of a midterm and final exam in introductory statistics courses should be revisited (Garfield, 1994).

Inefficient instructional methods will likely have the greatest effect on Watch-List and Off-The-Screen students.

Incoherence of the Statistics Curriculum

Another factor that may affect student learning is the incoherence in a statistics curriculum. This incoherence may take many forms. One common form in the typical two-semester sequence in introductory statistics is the use of different textbooks, instructors, and classroom practices. Unfortunately, and unbelievably, the practice of each instructor choosing a different textbook and employing vastly different instructional methods appears to be common. What is a student in a first-semester introductory statistics class to make of instructor 1 who uses textbook A and a traditional lecture style, along with a midterm and final exam that are the basis of assigning grades, and a second-semester course with instructor 2, textbook B, and a less traditional lecture style in which bi-weekly quizzes are used to assign final grades? The cumulative effect would seem to be to make it more difficult for some students to learn, particularly Watch-List and Off-the-Screen students. However, there is apparently no research documenting the effects of varied instructional practices in statistics like those mentioned above.

Another form of incoherence in a statistics curriculum occurs when recommended practices, conceptualizations, and guidelines applied in some statistics classes are ignored or even contradicted in others. One example is the emphasis that is placed on a priori power and sample size calculations in introductory courses. Too often, this emphasis appears for the single-factor ANOVA layout but is nowhere to be found for statistical tests that students have already learned (e.g., matched-pairs t-test, one- and two-sample binomial tests). A similar inconsistency sometimes occurs for confidence intervals.

An example of a more subtle kind of incoherence begins when recommended practices, such as the use of randomized designs and planned contrasts, discussion of internal validity, control of compounded Type I error rates, etc., are heavily emphasized in single- and multi-factor ANOVA-type designs in introductory courses. Students who proceed to a multiple regression course are less likely to have these practices discussed or recommended and who eventually sit for an HLM or SEM course are even less likely to have these ideas discussed. A reasonable question for a student to ask is: "If randomized designs, planned

contrasts, internal validity, and control of compounded Type I error rates are so important in introductory statistics classes, why are they not equally important in more advanced courses?" This kind of incoherence may contribute to poor statistical practice because students may fail to understand which concepts and recommended practices are always important.

It's The Software

The integration of data analysis software into statistics instruction has increased in the last 10 years, but it's implementation remains uneven. It is still possible for one student to sit for an introductory statistics course in which data analysis software is carefully integrated with every aspect of instruction (e.g., class demonstrations and exercises, homeworks, projects), and for another student to sit for an introductory statistics course in which no statistical software is used. Most worrisome is that these two students could be sitting for two sections of the same course offered by the same program in educational statistics.

Even when software has been incorporated into the instruction there are concerns. Dallal (1990) and Searle (1989) warned that the use of software in statistical instruction must be an integral part of the course, not just an add-on. Otherwise, the situation is ripe for misuse of statistical methods. Biehler (1993), Joyce and Showers (1982), and Cobb (1993) expressed similar concerns.

Perhaps the most striking feature of the factors described above is that they are directly or indirectly under an instructor's control. As discussed later, this suggests several remedies whose implementation is largely in the hands of educational statisticians.

What Factors Are Responsible for the Misuse of Statistical Procedures by Educational Researchers?

It's The Researcher

The magnitude of misuse of statistical procedures by educational researchers suggests inadequate statistical training. There is no other way to put it. As was the case with students, there is substantial heterogeneity in statistical backgrounds of many researchers; some have impressive statistical skills, while others are painfully lacking in even rudimentary skills. Many of the problems described for students may have applied to educational researchers when they were graduate students.

One way to recognize and perhaps sort out this heterogeneity is to characterize educational researchers (our faculty colleagues) by once again adapting the Elmore and Burney (1997) criteria:

• *With-the-Drill*

With-the-Drill researchers are, from the perspective of many educational statisticians, ideally suited to conduct statistical analyses. These researchers are highly motivated, detail-oriented, and show substantial self-discipline. They typically have had substantial graduate training in statistics. They value statistical analyses and pay careful attention to the assumptions underlying statistical procedures, demonstrate an appetite for new statistical procedures, and can be counted on to exercise care and caution in their statistical analyses.

• *Free-Agents*

Free-Agent researchers have usually had substantial graduate and/or postgraduate statistical training. They often value new statistical procedures and are quick to study and employ these techniques in their work. However, they are not particularily detail-oriented and sometimes pay too little attention to the assumptions underlying statistical procedures. These researchers are "free-agents" in the sense that educational statisticians are confident of their enthusiasm for statistical analyses and their skill in performing these analyses. As a result, they are given more leeway.

• *Watch-List*

Watch-List educational researchers are those who, while motivated, are often ill-prepared to credibly perform statistical analyses. They frequently have had little statistical training, and sometimes claim to be phobic about numbers. As a result, they often struggle in performing statistical analyses and correctly interpreting the results. They appear to be fully supportive of statistical analyses and are sometimes highly motivated, yet may still misuse statistical procedures. They rarely show an interest in learning more than the nuts-and-bolts of the procedures they need for their work. Educational statisticians probably spend more time with these researchers than those in any other category.

• *Off-the-Screen*

These are educational researchers who are poorly prepared, undisciplined, and unmotivated. They typically consider statistical analyses to be a minor part of their work, and may even resent the use of these procedures. Off-the-Screen educational researchers usually have had little graduate training in statistics, are genuinely disengaged from the statistical analysis, and can turn surly if the results of an analysis are not pleasing. Other things being equal, Free Agent and With-the-Drill researchers are least likely to misuse statistical procedures, whereas Watch-List and Off-the-Screen are the most likely.

It's the Software

Software-related concerns here are the same as those for students.

Pressure To Use Newest Statistical Methods And Produce Statistically Significant Findings

Another potential contributor to misuse begins with the tendency of some journals to publish papers employing newer statistical procedures. This places pressure on potential authors to use statistical procedures that they may not be familiar with, increasing the likelihood of misuse. Along the same lines, the pressure to produce statistically significant findings may play a role in the misuse of statistical procedures by educational researchers. In an academic setting, there is a view that the presence (or absence) of statistically significant results can affect publication of a research paper, which, in sufficient number, can eventually affect promotion and tenure as well as salary decisions. This pressure may result in some researchers misusing statistical procedures in the hope of finding statistical significance. All four categories of researchers are susceptible to this pressure.

Unlike the student case, factors responsible for the misuse of statistical procedures by our faculty colleagues are at best only indirectly under the control of educational statisticians.

What Remedies Are Needed to Reduce Misuse of Statistical Procedures in Educational Research?

Any remedy for the problems described above will, to some extent, need to be tailored to various types of students and educational researchers. These remedies are not exhaustive, but do share the characteristic that their implementation is largely under the control of educational statisticians in faculty roles.

Remedy #1 Put the Education Back Into Educational Statistics

Perhaps the most important remedy is for educational statisticians to put some "education" into their teaching as "statisticians." There is a substantial body of literature on effective ways to teach, and in particular on ways to effectively teach statistics. The recommendations of this literature for effective teaching and assessment practices need to be incorporated into statistics classes in ways that go beyond simply including realistic datasets in class examples or adding a few quizzes. Statistics instructors must meet the challenge of teaching heterogeneous classes by structuring the instruction in ways that promote learning. This includes using statistics textbooks that promote learning among students with vastly different learning styles and coordinating the instruction in ways that ensures consistency of ideas, practices, and recommendations across classes and instructors. Good papers here include Garfield (1993, 1995) and Giraud (1997). The Educational Statisticians Special Interest Group could contribute to this effort by setting aside a symposium slot at the AERA annual meeting for papers demonstrating effective teaching practices in statistics.

Educational statisticians must also convey the importance of careful statistical practice in the classroom as well as in other settings; for example, master's and dissertation committees, research projects, and consulting work. We must be unrelenting in our insistence on good statistical practice.

Finally, efforts to put the education back into educational statistics should include our faculty colleagues. It is unrealistic to expect a researcher with expertise in K-5 classroom reading instruction to demonstrate a comparable level of statistical expertise (or vice versa). Yet that expertise may be needed in analyzing data related to research in classroom reading instruction. One way that educational statisticians can help to reduce the misuse of statistical methods by our faculty colleagues is to offer instruction in particular topics (e.g., importance of model checking, introduction to HLM), perhaps under the guise of a seminar series. This effort needs to be carried out in a non-threatening and collegial manner. This can also be a good way to convince our colleagues of the value of additional statistical coursework for their students.

Remedy #2 Statistics Instructors Should Be Accountable For Their Teaching

Of course, faculty are indirectly accountable for the quality of their teaching through student and peer teaching evaluations. But a more important kind of accountability holds statistics instructors accountable for what students have (or have not) learned. Last fall Bill Harkness from the Penn State Department of Statistics gave a talk that focused on changes his department had made in their introductory undergraduate statistics course. The most interesting part of the talk was the systematic effort the department was making to talk to faculty across campus to ascertain the quality of the statistical work of students in subsequent classes. Where there was a consensus among faculty across campus that students were deficient in particular skills, the department moved to change the way those skills were taught and learned. Where students were missing needed skills, the instruction was changed to include these skills. This is, in effect, a validity study.

Statistics instructors and programs in educational statistics need to engage in this kind of accountability, that is, to find out what skills students are deficient in and to modify the statistics instruction accordingly. Educational statisticians in faculty roles are uniquely suited to this task because we frequently sit on the master's and dissertation committees of students who sat for our classes. The master/dissertation committee structure provides ample opportunity to evaluate a student's knowledge and the quality of their application of statistical procedures in ways that generate accountability evidence. Still, this effort cannot succeed with casual questions and cavalier responses; it requires hard questions and even harder answers. In some cases it may be necessary for an individual instructor or even an entire program to swallow their collective professional pride in the service of higher quality training of students in statistics.

Remedy #3 Insist on adequate training for students –programmatically or individually

Many educational statisticians would probably agree that this remedy is long overdue. Consider the following example: Suppose that in the course of serving on a dissertation committee it becomes clear that a student needs to perform a statistical analysis that they have had no training in, such as multiple regression or MANOVA. While there are always exceptions, the course of action that may do the most to minimize

statistical misuse (current and future) is to insist that the student sit for a course in multiple regression or MANOVA. This is a fundamentally simple and practical way to reduce the misuse of statistical procedures.

Remedy # 4 Change editorial policy for non-methodology journals such as RER and AERJ to have educational statisticians regularly serve as reviewers of quantitatively-oriented papers

This remedy would serve an important gate-keeping function that should increase the quality of statistical work reported in a journal.

Remedy # 5 Need to value research done in statistical education

One of the ongoing problems in educational statistics is that high quality research in statistical education does not appear to be particularly valued in promotion and tenure decision, annual salary raises, and even within professional associations. For some reason, there appears to be a widespread view that research in statistical education is not "statistical" enough, and in general represents a lower grade of work than many educational statisticians are willing to accept or participate in. This view persists even within the Educational Statisticians SIG—how many members of the SIG believe that high quality research in statistical education is badly needed and how many actually engage in this kind of research?

The failure to sufficiently value high quality research in statistical education is a destructive practice that undervalues an important body of work. Attaching an appropriate value and status to high quality work in statistical education would lead to more (and better) work in this area, which in turn would positively influence statistical instruction and help to reduce the misuse of statistical procedures.

Remedy #6 Need for a professional home for educational statisticians

Many disciplines are guided by their professional societies. These societies offer a mechanism for a profession to engage in discussion about issues of interest to the membership, and to offer solutions to long-standing problems. The spirit of this endeavor is that the issues we are confronted with in our daily activities, such as the misuse of statistical procedures, are common to those working in educational statistics. As such, the marshalling of the collective resources of professionals within a field can be an effective way to solve problems. This highly idealized description of a professional society may or may not hold for members

of the American Statistical Association, but it clearly does not hold for educational statisticians. This sort of professional vehicle for centralizing discussion is needed in educational statistics.

Future Research

There are many candidates for research activities whose goal is to increase the quality of statistical practice. There is a strong need for continued work on effective teaching and assessment practices in statistics, as well as the interplay between various instructional styles and student learning styles. There is also a need for validity studies of the effectiveness of current teaching practices in educational statistics, and studies of what constitutes adequate training in educational statistics. The continuing development of the statistical education literature provides a framework for reducing the misuse of statistical procedures.

References

Berkson, J. (1942). Tests of significance considered as evidence. *Journal of the American Statistical Association, 37*, 325-335.

Biehler, R. (1993). Software tools and mathematics education: The case of statistics. In *Learning from computers: Mathematics education and technology*. Eds W. Dorfler, C. Keitel, and K. Ruthven. Berlin: Springer.

Brewer, J.K. (1985). Behavioral statistics textbooks: Source of myths and misconceptions. *Journal of Educational Statistics, 10*, 252-268.

Brogan, D., & Kutner, M.H. (1986). Graduate statistics service courses. *The American Statistician, 40*, 252-254.

Bryk, A.S., Sebring, P.B., Kerbow, D., Rollow, S., & Easton, J.Q. (1998). *Charting Chicago school reform: Democratic localism as a lever for change*. Boulder, CO: WestviewPress.

Callaert, H. (1999). Nonparametric hypotheses for the two-sample location problem. *Journal of Statistics Education, Vol 7, No. 2.*

Cobb, G. (1987). Introductory textbooks: A framework for evaluation. *Journal of the American Statistical Association, 82*, 321-339.

Cobb, G. (1993). Reconsidering statistics education: A national science foundation conference. *Journal of Statistics Education, Vol 1, No. 1.*

Curtis, D.A., & Harwell, M. (1998). Training doctoral students in educational statistics in the United States: A national survey. *Journal of Statistics Education, Vol 1, No 1.*

Dallal, G.E. (1990). Statistical computing packages: Dare we abandon their teaching to others? *The American Statistician, 44*, 265-266.

Daniel, Larry G. (April, 1998). Statistical Significance Testing: A Historical Overview of Misuse and Misinterpretation with Implications for the Editorial Policies of Educational Journals. Paper presented at the annual meeting of the American Educational Research Association,

Draper, D. (1988). Rank-based robust analysis of linear models I: Exposition and review. *Statistical Science, 3*, 239-271.

Elmore, R.F., & Burney, D. (1997). School variation and systemic instructional improvement in Community School District 2, New York City. High Performance Learning Communities Project, University of Pittsburgh, Learning Research and Development Center.

Elmore, R. F. & Burney, D. (1999). Investing in teacher learning: Staff development and instructional improvement. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 263-291). San Francisco: Jossey-Bass.

Garfield, J. (1993). Teaching Statistics Using Small-Group Cooperative Learning. *Journal of Statistics Education, Vol 1, No 1.*

Garfield, J. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education, Vol 2, No 1.*

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*, 25-34.

Giraud, G. (1997). Cooperative Learning and Statistics Instruction. *Journal of Statistics Education, Vol 5, No 3.*

Glisson, C. A., & Hudson, W.W. (1981). Applied Statistical Misuse in Educational Research: An Admissions Criteria Example. *Journal of Education for Social Work; 17*, 35-44.

Harwell, M.R., & Gatti, G.G. Rescaling ordinal data into interval data in educational research. (*in press*)

Harwell, M.R., Herrick, M.L., Curtis, D., Mundfrom, D., & Gold, K. (1996). Evaluating statistics texts used in education. *Journal of Educational and Behavioral Statistics, 21*, 3-34.

Helberg, C. (1996). Pitfalls of Data Analysis. ERIC Clearinghouse on Assessment and Evaluation (ED410231).

Hogg, R. V. (1990), "Statisticians Gather to Discuss Statistical Education," *Amstat News, 169*, 19-20.

----- (1991), "Statistical Education: Improvements Are Badly Needed," *The American Statistician, 45*, 342-343.

Hooke, R. (1980). Getting people to use statistics properly. *The American Statistician, 34*, 39-42.

Keselman, H.J., Huberty, C.J., Lix, L.M., Olejnik, S., Cribbie, R.A., Donohue, B., Kowalchuk, R.K., Lowman, L.L., Petosky, M.D., Keselman, J.C., & Levin, J.R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*, 350-386.

Knapp, M.S., & McLaughlin, M. (April, 1999). *The district role in the improvement of teaching: Framing the conversation and research.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Lykken, D.T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151-159.

Magel, R.C. (1998). Using Cooperative Learning in a Large Introductory Statistics Class. *Journal of Statistics Education, Vol 6, No 3.*

Neter, J., Kutner, M. H., Nachtsheim, C.J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed). Boston: McGraw Hill.

Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology, 46*, 128-148.

Searle, S.R. (1989). Statistical computing packages: Some words of caution. *The American Statistician, 43*, 189-190.

Singer, J. D., & Willett, J.B. (1990). Improving the teaching of applied statistics: Putting the data back into data analysis. *The American Statistician, 44*, 223-230.

Snee, R. D. (1993). What's Missing in Statistical Education? *The American Statistician, 47*, 149- 154.

Wilson, J. (1973). Three myths in educational research. *Educational Research, 16*, 17-19.